

Evaluating the reliability of microsatellite genotyping from low-quality DNA templates with a polynomial distribution model

HE Gang^{1†}, HUANG Kang^{1†}, GUO SongTao^{1*}, JI WeiHong², QI XiaoGuang¹, REN Yi³, JIN XueLin⁴ & LI BaoGuo^{1,3*}

¹ Key Laboratory of Resource Biology and Biotechnology in Western China of Ministry of Education, College of Life Sciences, Northwest University, Xi'an 710069, China;

² Ecology & Conservation Group, Institute for Natural Sciences, Massey University, Auckland 904, New Zealand;

³ Institute of Zoology, Shaanxi Academy of Sciences, Xi'an 710032, China;

⁴ Shaanxi Wild Animal Rescue and Research Center, Xi'an 710402, China

Received March 2, 2011; accepted June 20, 2011

Molecular studies using trace DNA, such as from museum specimens, ancient or forensic samples and samples obtained noninvasively, often have a common problem of low quality of DNA templates. Amplification errors, such as allelic dropout and false allele, may arise during polymerase chain reaction (PCR) using such samples. A mathematical model which treats homozygotes and heterozygotes discriminately has been developed to measure sample quality and compute the confidence level of using multiple-tube approaches. We use plucked hair samples collected from 26 individual Sichuan snub-nosed monkeys (*Rhinopithecus roxellana*) to test the model. In this case, a confidence level of 99% can be achieved by three positive PCRs. If the sample quality is very poor and requires many PCR replicates, an alternative multiple-step genotyping method is recommended. This model enables researchers to optimize experimental protocols through pilot studies and obtain reliable genetic information using noninvasive sampling method.

Sichuan snub-nosed monkey, noninvasive sampling, microsatellite, genotyping errors, polynomial distribution model

Citation: He G, Huang K, Guo S T, et al. Evaluating the reliability of microsatellite genotyping from low-quality DNA templates with a polynomial distribution model. Chinese Sci Bull, 2011, 56: 2523–2530, doi: 10.1007/s11434-011-4634-5

Noninvasive sampling has become an important method for studying endangered and cryptic species [1]. It allows for genetic analysis of wild animals without the need for handling and causing stress [2]. However, DNA templates from noninvasive samples [3–5], museum specimens [6,7], ancient DNA [8] and forensic samples [9] are often degraded or contaminated. The amount of target DNA available for genetic typing can be very low and is often in the picogram range [9–12]. When such low quality DNA is used for microsatellite profiling, two main genotyping errors can occur leading to spurious results: (i) false homozygote, also known as allelic dropout, which is misidentification of a

heterozygote as a homozygote because of the recognition of only one of the two alleles [10,13,14]; and (ii) false allele, which is the amplification of a polymerase chain reaction (PCR) artifact [10,15,16].

To reduce the chance of false results, Navidi et al. [13] suggested a multiple-tube approach, in which the DNA is distributed among several tubes. Compared to a single tube approach, it provides more reliable genotyping from diluted DNA samples. Taberlet et al. [10] improved this method, and for homozygous individuals, a confidence level of 99% in genotyping assignments could be reached by analyzing seven independent PCRs. Miller et al. [17] developed a maximum-likelihood approach to assess genotype reliability and reduce the number of re-amplifications for heterozygous individuals. When allelic dropout rates are low, the

† These authors contributed equally to the work.

* Corresponding authors (email: songtaoguo@nww.edu.cn; baoguoli@nww.edu.cn)

reduction in the number of PCR replicates is typically 40%–50% [17]. Frantz et al. [18] proposed a modification of the multiple-tube approach that may reduce the number of amplifications required to obtain reliable genotypes. Consensus genotypes are determined after two initial positive PCRs for heterozygotes and three for homozygotes [18,19]. In general, the lower the amount of template in the PCR, the more prone it is to genotyping errors and the more PCR replicates are required [20]. Researchers have attempted to minimize the number of PCR replications by quantifying the amount of amplifiable DNA present in extracts [11,21] or by calculating a quality index for each sample [22].

The use of multiple PCR replicates can result in the sample becoming exhausted if many loci are analyzed, incurring additional costs [20,23]. Efficiently evaluating the reliability of microsatellite genotyping using low quality DNA templates can help optimize the experimental protocol. In this study, we aim to answer the following questions for a given batch of samples: (i) how many amplifications should be performed to reach a given confidence level? (ii) how should the confidence level of a protocol be calculated?

1 A polynomial distribution model and model simulation

1.1 A polynomial distribution model of amplification errors

To answer the above-mentioned questions, we have developed a mathematical model to compute the confidence level of multiple-tube approaches. The PCR of homozygous and the heterozygous loci generate different errors, with mainly false alleles at homozygous loci and false homozygotes at heterozygous loci. Because mechanisms of these two types of errors are different, they should be discriminated.

We consider an amplification that provides unambiguous target sequences in a polyacrylamide gel as a positive PCR, otherwise we consider it as a negative PCR. If more than two alleles are detected in an amplification, we also treat it as a negative PCR. Performing n positive PCRs for each individual at each locus, we record any allele that appears at least m times ($1 \leq m \leq n$). Figure 1 is a schematic diagram that shows the genotyping approach with $n=5$ and $m=2$.

If a sample consistently produces negative PCRs, we

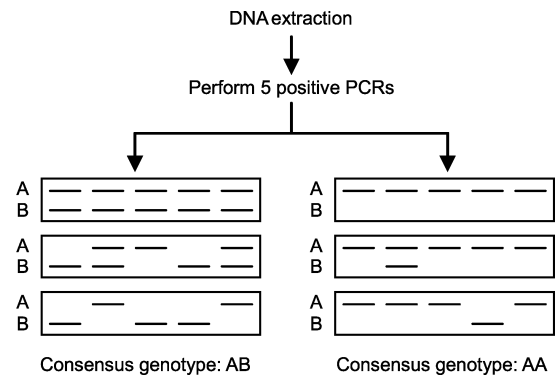


Figure 1 Schematic diagram of multiple-tube genotyping approach with $n=5$ and $m=2$. There are three polyacrylamide gels at each side and each gel has five independent positive PCRs of the same individual. If an allele appears at least twice in three positive PCRs, it is regarded as the consensus genotype.

redigest the sample or abandon the genotyping. Only positive PCRs are recorded, because negative PCRs can be a result of not only low template concentration but also human errors and machine irregularities, while PCR-generated errors in positive PCRs can only be caused by low DNA quality. A positive PCR for a heterozygous locus can have three possible interpretations: a correct heterozygote result, or two types of false homozygotes. A positive PCR for homozygous locus can also have two possible interpretations: correct or false allele.

Let R_1 be the correct rate of heterozygous individual in positive PCRs and R_2 be that of homozygotes in positive PCRs. R_1 and R_2 can be the measurements of the samples quality under given experimental conditions. Table 1 shows each result and the estimated probabilities of amplification of heterozygotes and homozygotes.

Let E_1 be the event that the two alleles of heterozygotes both appear at least m times in n positive PCRs ($1 \leq m \leq n$). If the two alleles are labeled as A and B, the results can be summarized as the number of times in which AA, AB and BB appeared: i, j and k . The probability of E_1 is calculated by a trinomial distribution function as follows:

$$P(E_1) = \sum_{\substack{i+j+k=n \\ i+j \geq m \\ j+k \geq m}} \frac{n!}{i!j!k!} \left(\frac{1-R_1}{2} \right)^{i+k} R_1^j. \quad (1)$$

Let E_2 be the event that a false allele appears at most $m-1$

Table 1 Results of heterozygote and homozygote and their estimate probability in a single positive PCR

Heterozygote (AB)'s amplification result	Estimate probability	Homozygote (AA)'s amplification result	Estimate probability
Correct (AB)	R_1	Correct (AA)	R_2
False homozygote (AA)	$\frac{1}{2}(1-R_1)$	False allele (AB or etc.)	$1-R_2$
False homozygote (BB)	$\frac{1}{2}(1-R_1)$		

times in n positive PCRs. The probability of E_2 can be expressed by a binomial cumulative distribution function as follows:

$$P(E_2) = \sum_{i=0}^{m-1} \frac{n!}{i!(n-i)!} (1-R_2)^i R_2^{n-i}. \quad (2)$$

The correct genotyping rate P of a locus with a heterozygosity of H is given by

$$P = HP(E_1) + (1-H)P(E_2), \quad (3)$$

where P ranges from $P(E_1)$ to $P(E_2)$.

This model has three assumptions: (i) the DNA quality is equal across all samples; (ii) both alleles of a heterozygote are equally likely to dropout; and (iii) each homozygous individual generates only one type of false allele.

1.2 Model simulation

We calculated a set of minimum threshold of sample quality for multiple-tube genotyping approach at the 99% confidence level. The results (Table 2, Figure 2) indicate that the minimum threshold of R_1 decreases and the minimum threshold of R_2 increases as the number of amplification repeats increases when m is fixed. Note that after eight positive PCRs with $m = 1$, the minimum threshold of R_1 becomes zero, which indicates that if the template concentration is extremely low, each positive PCR can only detect one allele of a heterozygote. Eight positive PCRs ensure each allele appears at least once at a probability of 99%. Similarly, 12 and 15 positive PCRs ensure each allele appears at least twice and three times, respectively.

We used hair samples from *Rhinopithecus roxellana*, a

Table 2 Sample quality threshold in multiple-tube approaches at the 99% confidence level

n positive PCRs	R_1 minimum	R_2 minimum	n positive PCRs	R_1 minimum	R_2 minimum
$m = 1$					
1	0.9900	0.9900	10	0.0886	0.9845
2	0.8586	0.9950	11	0.0171	0.9859
3	0.6580	0.9967	12	0	0.9872
4	0.4682	0.9975	$m = 3$		
5	0.3069	0.9980	3	0.9967	0.7846
6	0.1730	0.9983	4	0.9411	0.8591
7	0.0618	0.9986	5	0.8343	0.8944
8	0	0.9987	6	0.7128	0.9153
$m = 2$			7	0.5941	0.9292
2	0.9950	0.9000	8	0.4843	0.9392
3	0.9172	0.9411	9	0.3852	0.9467
4	0.7782	0.9580	10	0.2964	0.9525
5	0.6298	0.9673	11	0.2170	0.9572
6	0.4920	0.9732	12	0.1459	0.9610
7	0.3698	0.9773	13	0.0821	0.9642
8	0.2630	0.9803	14	0.0246	0.9669
9	0.1699	0.9826	15	0	0.9693

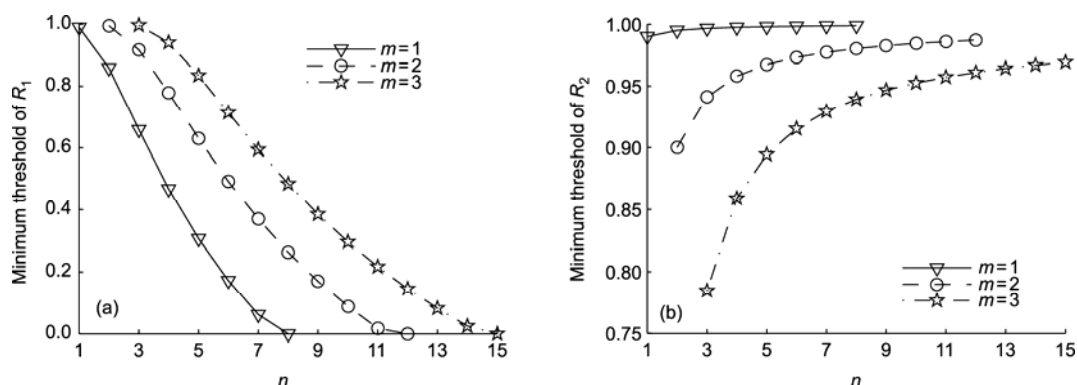


Figure 2 Minimum threshold of R_1 (a) and R_2 (b) at the 99% confidence level using multiple-tube approaches.

colobine monkey endemic to China, to demonstrate the use of this model. Because of poaching and habitat fragmentation, this species has become endangered and is included in the IUCN Red List of Threatened Species [24]. Their distribution is now restricted to the mountains of central China [25]. Genetic analysis using fecal and hair samples have been carried out to investigate ecological and behavioral aspects of this species [26–28]. Genetic markers previously used include mtDNA and microsatellite markers [29,30]. Genotyping using fecal and hair samples were subject to the problems of low quality DNA templates and genotyping errors.

2 Materials and methods

Plucked hair samples were collected from 15 individuals of *R. roxellana* from Shaanxi Province Rare Wildlife Rescue and Breeding Center (Lou-Guan-Tai) and 11 individuals from a wild group in Qinling Mountains (Zhouzhi National Nature Reserve). We collected 10–20 hairs from each individual. Hair samples were then stored at room temperature in silica gel for drying and DNA was extracted within one year. All research protocols were reviewed and approved by Chinese Academy of Sciences and Shaanxi Zoological Society that regulates research with animals, and that the research complied with the protocols approved by institutional animal care committees, and adhered to the legal requirements of China.

2.1 DNA extraction

DNA was extracted from the hair samples using a proteinase K digestion in a PCR compatible buffer [31]. Five hairs can provide template for 100 PCRs, and procedures are as follows:

- (1) Rinse five hairs in 70% ethanol and then in sterile distilled water in a sterile 0.2 mL EP tube.
- (2) Dry the hairs in a drying oven for 20 min at 50°C. Cut hair roots approximately 5 mm from the end of the root with a sterile scissor and transfer to another sterile EP tube. Centrifuge at 4000 r/min for 1 min.
- (3) Add 20 μL of follicle lysis buffer (MgCl₂ 0.2 mmol/L,

Tris · HCl 10 mmol/L (pH 8), KCl 50 mmol/L, proteinase K 0.2 mg/mL), incubate at 65°C for 30 min and denature the enzyme at 95°C for 15 min in a thermocycler (Eppendorf Mastercycler Gradient, Germany).

- (4) Centrifuge at 12000 r/min for 3 min. Transfer the supernatant into another sterile EP tube and dilute to 100 μL. Store at 4°C.

2.2 Primer design

We selected 19 polymorphic microsatellite loci from related species of *R. roxellana*, 10 loci from *R. bieti* [30]; 4 from *Macaca fuscata* [32] and 5 from *Macaca mulatta* [33]. Primers were synthesized by Sangon Biotech (Shanghai) Co., Ltd.

After genotyping and screening, we selected five high polymorphic loci for this study. Their primer information is shown in Table 3. Among other 14 microsatellite loci, 2 have very low degree of polymorphism, 7 are monomorphic and 5 cannot be amplified in *R. roxellana*.

2.3 Amplification and genotyping

PCR amplifications were performed with a thermocycler (Eppendorf Mastercycler gradient) in 10 μL reactions containing 1 μL of template solution, 10 mmol/L Tris · HCl, 50 mmol/L KCl, 1.5 mmol/L MgCl₂, 250 μmol/L each of 4 dNTPs, 0.2 μmol/L of each primer and 0.25 U golden DNA polymerase (Tiangen, China). The PCR program consisted of an initial denaturation at 95°C for 4 min followed by 35 cycles at 95°C for 30 s; 60°C for 45 s and 72°C for 45 s; and final extension at 72°C for 10 min. Amplification products were separated by electrophoresis on a 10% polyacrylamide gel at 6 V/cm for 12 h and visualized after silver staining. We perform five positive PCRs for each individual at each locus, and record any allele observed in at least three amplifications (*m* = 3).

3 Results

Three individuals at locus *D6s493* and one individual at

Table 3 Profile of STR loci

Locus	Size (bp)	Array	Primer (5'→3')	T _a (°C)	No. of alleles
<i>D20s206</i>	~130	TATC	Fwd: TCCATTATTCCCCTCAAACA Rev: GGTTTGCCATTCAGTTGAGA	52	5
<i>D5s1457</i>	~150	TAGA	Fwd: TAGGTTCTGGGCATGTCTGT Rev: TGCTTGGCACACTTCAGG	57	5
<i>D6s493</i>	~260	TATC	Fwd: ATCCCAACTCTTAAATGGGC Rev: TTCCATGGCAGAAATTGTTT	53	6
<i>D7s2204</i>	~260	GATA	Fwd: TCATGACAAAACAGAAATTAAGTG Rev: AGTAAATGGAATTGCTTGTTACC	54	7
<i>D14s306</i>	~190	TATC	Fwd: AAAGCTACATCCAAATTAGGTAGG Rev: TGACAAAGAACTAAAATGTCCC	52	6

locus *D14s306* consistently produced negative PCRs and genotypes of these individuals were not obtained. In total, we obtained 630 positive PCRs. Results are shown in Table 4.

Among the 630 positive PCRs, false homozygote occurred 9 times and false allele occurred 4 times. Genotyping results are shown in Table 5, through which we obtained $\hat{R}_1 = 0.9739$ (one-sided 95% CI = [0.9567, 1]) and $\hat{R}_2 = 0.9860$ (one-sided 95% CI = [0.9706, 1]). According to eq. (3), a confidence level of 99% can be achieved by performing three positive PCRs and genotyping with $m = 2$. The correct genotyping rate ranges from 0.9990 to 0.9994.

4 Discussion

4.1 Amplification error

Compared with blood and tissue, the concentration of DNA extracted from a hair follicle or other noninvasive sample is low, which is likely to lead to a variety of errors in PCR amplification, therefore reducing the reliability of genotyping. We encountered two types of amplification errors during this study, false alleles and allele dropouts (false homozygote).

The false homozygotes may due to random sampling of template DNA in the very diluted extract [10] and imbalance in the efficiency of the amplification for each allele, especially when the sizes of the alleles are significantly different [23]. In this study, R_1 expresses the correct rate of heterozygotes in positive PCRs. R_1 less than 0.9172 indicates the samples are in poor quality, in which case more than three positive PCRs should be performed.

Table 4 Genotyping results at 5 microsatellite loci

Locus	Heterozygote	Homozygote	Unknown
<i>D20s206</i>	13	13	0
<i>D5s1457</i>	11	15	0
<i>D6s493</i>	19	4	3
<i>D7s2204</i>	14	12	0
<i>D14s306</i>	12	13	1

Table 5 Summary of PCR errors

Locus	No. of positive PCRs	No. of PCR results consistent to consensus genotype	Rate of false allele	Rate of false homozygote
<i>D20s206</i>	130	126	0/65	4/65
<i>D5s1457</i>	130	129	1/75	0/55
<i>D6s493</i>	115	115	0/20	0/95
<i>D7s2204</i>	130	127	1/60	2/70
<i>D14s306</i>	125	120	2/65	3/60
Total	630	617	4/285	9/345

The false allele is probably caused by slippage mutations that are insertions or deletions of one or more repeats in a microsatellite locus [15,34]. If a slippage mutation occurs during the first few cycles of the amplification, the artificial allele can be detected at the same intensity as true alleles only when the number of target molecules is very low [10,23]. Previous studies have shown that the slippage rate is correlated to length and repeat motif for long microsatellites [35–37]. Short ones are far less known but the issue of a minimum threshold length for DNA slippage remains contentious [38]. We use R_2 to express the rate of correct homozygote amplification in positive PCRs, and \hat{R}_2 is larger than 0.98 in many studies (e.g. [20,39]).

4.2 Model assumptions

The first assumption is that samples have similar initial template concentrations. When this assumption is met, each heterozygote at a locus is equally likely to incur an allelic dropout, and each homozygote at a locus is equally likely to incur a false allele.

The second assumption is that alleles at the same microsatellite locus are positioned at the same location in the genome and their content and context are similar. Their efficiencies of amplification are approximately equal unless the sizes of alleles are dramatically different. Differences in amplification efficiency of alleles of different size were supported by some studies [3,18]. However Constable et al. [40] did not detect such a difference.

By contrast, if one allele in heterozygote is much longer than the other, which makes their efficiencies of amplification significantly different, then eq. (1) will be transformed as (see Appendix 1 for derivation):

$$P(E_1) = \sum_{i=m}^n \frac{n!}{i!(n-i)!} R_1^i (1-R_1)^{n-i}. \quad (4)$$

The third assumption is conservative because a slippage mutation can insert or delete one or more repeats in a microsatellite locus [15,34]. In such a case, there can be more than one type of false allele of each homozygote. However, details such as the ratio of the occurrence of false alleles and the minimum threshold length for DNA slippage are unknown [38]. This assumption caused our model to be strict when n is high. However, for low numbers of positive PCRs, the minimum threshold of R_2 is in a practical range.

4.3 Microsatellite genotyping strategy

In molecular biology studies, pilot experiments are usually carried out to test the conditions and protocols. Even so, it is difficult to ensure every experiment is successful. In microsatellite genotyping using low quality DNA with a large sample size, we recommend researchers evaluate sample quality before the experiments by taking five positive PCRs

at each locus of several randomly selected individuals. This should be followed by genotyping with $m=3$ to obtain parameters \hat{R}_1 and \hat{R}_2 because the minimum threshold of R_1 and R_2 are both relatively low ($m=3$, $n=5$). If $\hat{R}_1 < 0.8344$ or $\hat{R}_2 < 0.8944$, the pilot experiments are unreliable, and a different genotyping approach should be chosen to obtain more accurate values (re-genotyping the existing amplifications or performing additional experiments), or sample quality and experimental procedures should be improved.

When the pilot experiments are reliable, an appropriate approach can be chosen from Table 2. The total number of positive PCRs of heterozygotes and homozygotes should make the lower bound of the one-sided confidence interval (CI) of R_1 and R_2 larger than the minimum threshold of approach to be selected. The one-sided confidence interval was calculated by the Jeffery method. Let x be the number of successes in a random sample of size n . The lower limit of $100(1-\alpha)\%$ confidence interval is defined by $CI_1^L = [B_{\alpha, x+1/2, n-x+1/2}, 1]$, where $B_{\alpha, x+1/2, n-x+1/2}$ is the α quantile of a beta distribution Beta ($x+1/2, n-x+1/2$) [41].

For good quality samples ($\hat{R}_1 > 0.9172$, $\hat{R}_2 > 0.9411$), three positive PCR should be performed and genotyped with

$m=2$. If the first two positive PCRs generate equal genotypes, the third one is not necessary because it will not affect the consensus genotype. For poor quality samples, usually with a low false allele rate but a high allelic dropout rate, the selection approach ($m=m_s$, $n=n_s$) requires too many replicates. An alternative multiple-step method modified from the method of Taberlet et al. [10] which is able to minimize the number of replicates to be used, involves performing m_s initial replicates, genotyping with $m=m_s$, and accepting the genotype if it is heterozygous. Otherwise one or two additional positive PCRs (single addition method, see [17]) should be performed, genotyped with $m=m_s$ and heterozygous genotypes accepted. This step should be repeated until the rest of the amplifications do not alter the consensus genotype or n is up to n_s before accepting the obtained genotype.

We simulated the estimated number of replicates by calculating the sum of the products of the probability and the number of replicates of each event. The least n , followed by the least m , should be selected. The program for this calculation can be found in Program S1. Results are shown in Figure 3. When $R_1 > 0.5$ and $R_2 > 0.8$, this method can reduce the number of replicates by one third. When $R_1 > 0.9$ and $R_2 > 0.95$, this method can reduce the number by 28% of replicates ($H=0.5$). In the samples of our example, the

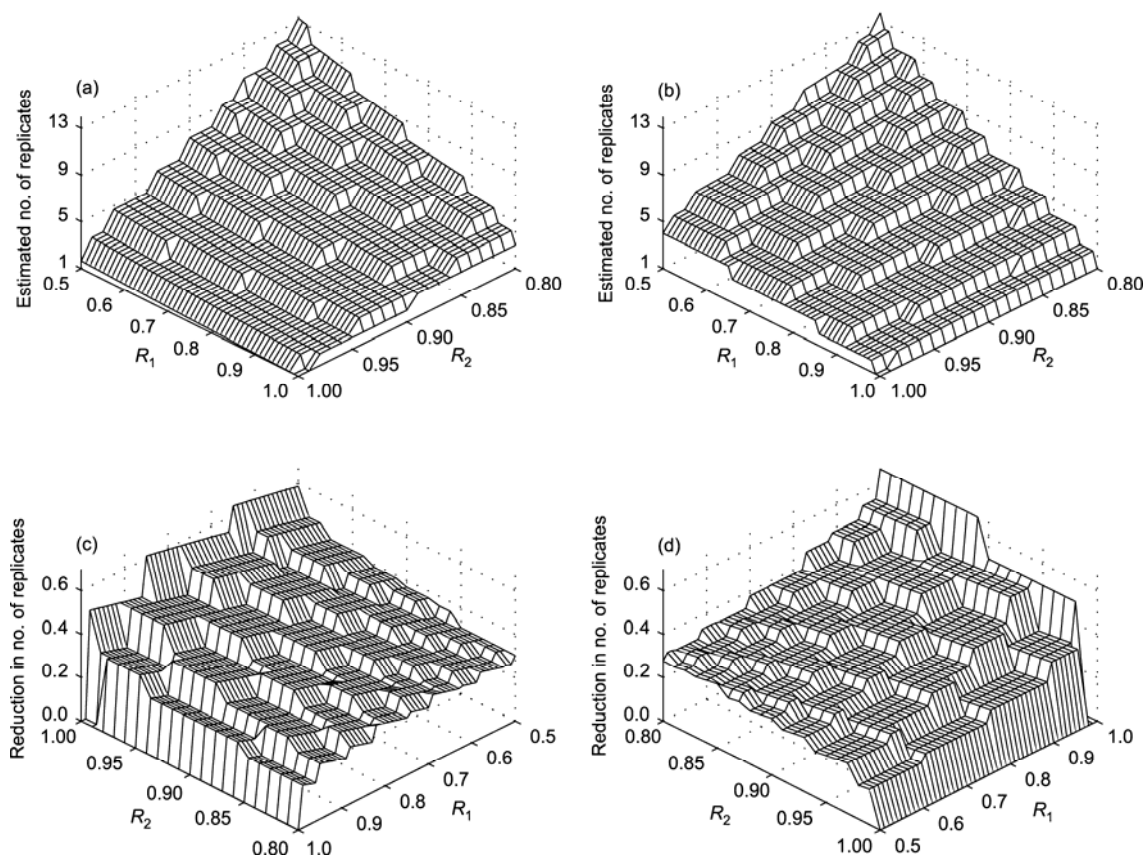


Figure 3 The estimated number of replicates and the reduction in the number of replicates of heterozygotes (a), (c); and homozygotes (b), (d), using the multiple-steps method at the 99% confidence level.

estimated number of replicates for homozygotes and heterozygotes were 2.0512 and 2.0276, respectively.

If the samples are poor quality and none of the approaches in Table 2 are practical, other measures should be taken. These include improving sampling and extraction procedures and optimizing the amplification conditions (e.g., increasing the template concentration or using a better *Taq* enzyme) or abandoning the genotyping. Compared to optimizing the amplification conditions, improving the sample quality is more realistic.

Our method allowed a number of improvements over those of similar studies (e.g. [10,17,18]). Specifically, (i) we evaluated the sample quality during pilot experiments and propose that the appropriate genotyping approach can be selected from Table 2; (ii) this model considers the effect of false alleles and m is variable; (iii) if the chosen approach needs too many positive PCRs, a researcher can use an alternative multiple-step method to reduce the number of replicates. This can minimize the number of replicates for heterozygotes and homozygotes. A computer program, Reliability Calculator, implementing this model, is available in Program S1.

We thank Zhouzhi National Nature Reserve for giving us permission to conduct this study. We thank Northwest University students W Wei and H T Zhao for their field work assistance and L L Wu and H Y Zhang for their assistance with laboratory experiments, and Prof. S W Chang and Prof. Y G He for providing comments on the manuscript. This work was supported by the National Natural Science Foundation of China (30970379, 30970444 and 30970168), the Natural Science Foundation of Shaanxi Province (2009JQ3001), the Scientific Research Foundation of the Education Department of Shaanxi Province (09JK748), and the Opening Foundation of the Key Laboratory of Resource Biology and Biotechnology in Western China (Northwest University) of Ministry of Education.

- Woodruff D S. Non-invasive genotyping of primates. *Primates*, 1993, 34: 333–346
- Taberlet P, Luikart G. Noninvasive genetic sampling and individual identification. *Biol J Linn Soc*, 1999, 68: 41–55
- Gagneux P, Boesch C, Woodruff D S. Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Mol Ecol*, 1997, 6: 861–868
- Bradley B J, Boesch C, Vigilant L. Identification and redesign of human microsatellite markers for genotyping wild chimpanzee (*Pan troglodytes verus*) and gorilla (*Gorilla gorilla gorilla*) DNA from faeces. *Conserv Genet*, 2000, 1: 289–292
- Regnaut S, Lucas F S, Fumagalli L. DNA degradation in avian faecal samples and feasibility of non-invasive genetic studies of threatened capercaillie populations. *Conserv Genet*, 2006, 7: 449–453
- Hajibabaei M, Smith M, Janzen D H, et al. A minimalist barcode can identify a specimen whose DNA is degraded. *Mol Ecol Notes*, 2006, 6: 959–964
- Sefc K M, Payne R B, Sorenson M D. Single base errors in PCR products from avian museum specimens and their effect on estimates of historical genetic diversity. *Conserv Genet*, 2007, 8: 879–884
- Cooper A, Wayne R. New uses for old DNA. *Curr Opin Biotech*, 1998, 9: 49–53
- Park M J, Lee H Y, Chung U, et al. Y-STR analysis of degraded DNA using reduced-size amplicons. *Int J Legal Med*, 2007, 121: 152–157
- Taberlet P, Griffin S, Goossens B, et al. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res*, 1996, 24: 3189–3194
- Morin P A, Chambers K E, Boesch C, et al. Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Mol Ecol*, 2001, 10: 1835–1844
- Alonso A, Martín P, Albarrán C, et al. Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies. *Forensic Sci Int*, 2004, 139: 141–149
- Navidi W, Arnheim N, Waterman M S. A multiple-tube approach for accurate genotyping of very small DNA samples by using PCR: Statistical considerations. *Am J Hum Genet*, 1992, 50: 347–359
- Fernando P, Vidya T N C, Rajapakse C, et al. Reliable noninvasive genotyping: Fantasy or reality? *J Hered*, 2003, 94: 115–123
- Schlötterer C, Tautz D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res*, 1992, 20: 211–215
- Foucault F, Praz F, Jaulin C, et al. Experimental limits of PCR analysis of (CA)_n repeat alterations. *Trends Genet*, 1996, 12: 450–452
- Miller C R, Joyce P, Waits L P. Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics*, 2002, 160: 357–366
- Frantz A C, Pope L C, Carpenter P J, et al. Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Mol Ecol*, 2003, 12: 1649–1661
- Hansen H, Ben-David M, McDonald D B. Effects of genotyping protocols on success and errors in identifying individual river otters (*Lontra canadensis*) from their feces. *Mol Ecol Resour*, 2008, 8: 282–289
- Arandjelovic M, Guschanski K, Schubert G, et al. Two-step multiplex polymerase chain reaction improves the speed and accuracy of genotyping using DNA from noninvasive and museum samples. *Mol Ecol Resour*, 2009, 9: 28–36
- Ball M C, Pither R, Manseau M, et al. Characterization of target nuclear DNA from faeces reduces technical issues associated with the assumption of low-quality and quantity template. *Conserv Genet*, 2007, 8: 577–586
- Miquel C, Bellemain E, Poillot C, et al. Quality indexes to assess the reliability of genotypes in studies using noninvasive sampling and multiple-tube approach. *Mol Ecol Notes*, 2006, 6: 985–988
- Goossens B, Waits L P, Taberlet P. Plucked hair samples as a source of DNA: Reliability of dinucleotide microsatellite genotyping. *Mol Ecol*, 1998, 7: 1237–1241
- IUCN. IUCN Red List of Threatened Species. Version 2010.4. 2010. <http://www.iucnredlist.org>
- Li B G, He P J, Yang X Z, et al. The present status of the Sichuan snub-nosed monkey in the Qinling Mountains of China, and a proposed conservation strategy for the species. *Biosphere Conserv*, 2001, 3: 107–114
- Pan D, Li Y, Hu H X, et al. Microsatellite polymorphisms of Sichuan golden monkeys. *Chinese Sci Bull*, 2005, 50: 2850–2855
- He L, Zhang Y G, Li D Q, et al. Analysis on mitochondrial DNA D-loop sequences genetic polymorphism of *Rhinopithecus roxellana* (in Chinese). *Chin J Zool*, 2010, 1: 70–76
- Guo S T, Ji W H, Li M, et al. The mating system of the Sichuan snub-nosed monkey (*Rhinopithecus roxellana*). *Am J Primatol*, 2010, 72: 25–32
- Guo S T. Inbreeding avoidance, paternity exclusion and mating system of Sichuan snub-nosed monkey in Qinling, China (in Chinese). Doctor Dissertation. Xi'an: Northwest University, 2007
- Liu Z J, Ren B P, Hao Y L, et al. Identification of 13 human microsatellite markers via cross-species amplification of fecal samples from *Rhinopithecus bieti*. *Int J Primatol*, 2008, 29: 265–272
- Allen M, Engström A S, Meyers S, et al. Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: Sensitivity and matching probabilities. *J Forensic Sci*, 1998, 43: 453–464
- Zhang H, Li J H, Zhao J Y, et al. Morphological characters and genetic polymorphism analysis by microsatellite loci in Rhesus monkey stock from Wannan Mountains (in Chinese). *Lab Anim Comparative Med*, 2008, 28: 225–229

- 33 Rogers J, Garcia R, Shelledy W, et al. An initial genetic linkage map of the rhesus macaque (*Macaca mulatta*) genome using human microsatellite loci. *Genomics*, 2006, 87: 30–38
- 34 Viguera E, Canceill D, Ehrlich S D. Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J*, 2001, 20: 2587–2595
- 35 Primmer C R, Ellegren H. Patterns of molecular evolution in avian microsatellites. *Mol Biol Evol*, 1998, 15: 997–1008
- 36 Whittaker J C, Harbord R M, Boxall N, et al. Likelihood-based estimation of microsatellite mutation rates. *Genetics*, 2003, 164: 781–787
- 37 Sainudiin R, Durrett R T, Aquadro C F, et al. Microsatellite mutation models, insights from a comparison of humans and chimpanzees. *Genetics*, 2004, 168: 383–395
- 38 Leclercq S, Rivals E, Jarne P. DNA slippage occurs at microsatellite loci without minimal threshold length in humans: A comparative genomic approach. *Genome Biol Evol*, 2010, 2: 325–335
- 39 Prugh L R, Ritland C E, Arthur S M, et al. Monitoring coyote population dynamics by genotyping faeces. *Mol Ecol*, 2005, 14: 1585–1596
- 40 Constable J L, Ashley M V, Goodall J, et al. Noninvasive paternity assignment in Gombe chimpanzees. *Mol Ecol*, 2001, 10: 1279–1300
- 41 Cai T T. One-sided confidence intervals in discrete distributions. *J Stat Plann Infer*, 2005, 131: 63–88

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Supporting Information

- Appendix 1** The derivation of eq. (4)
Program S1 Reliability Calculator

The supporting information is available online at csb.scichina.com and www.springerlink.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.